PCT WORLD INTELLEC International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/00, C12N 15/00
C12Q 1/68

A1

(11) International Publication Number: WO 92/15602

(43) International Publication Date: 17 September 1992 (17.09.92)

(21) International Application Number: PCT/US92/01624

(22) International Filing Date: 28 February 1992 (28.02.92)

(30) Priority data:
662,216 28 February 1991 (28.02.91) US

(71) Applicant: DANA-FARBER CANCER INSTITUTE, INC. [US/US]; 44 Binney Street, Boston, MA 02115 (US).

(72) Inventor: SAGER, Ruth ; 30 Codman Road, Brookline, MA 02146 (US).

(74) Agent: FRASER, Janis, K.; Fish & Richardson, 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: CANCER DIAGNOSIS AND THERAPY

(57) Abstract

A method for determining the presence of cancerous cells in a tissue from a patient, which method includes the steps of providing either (a) a nucleic acid probe including a nucleotide sequence at least 8 nucleotides in length which is identical to a portion or all of the coding sequence of a candidate tumor suppressor gene, or (b) an antibody specific for a candidate tumor suppressor gene product; obtaining from a patient a first tissue sample potentially including cancerous cells; providing a second tissue sample, substantially all of the cells of which are non-cancerous; and comparing, by use of either the probe or the antibody, the levels of expression of the candidate tumor suppressor gene in the first and second tissue sample, wherein an amount of hybridization or immune complex formation, as the case may be, in the first tissue sample less than one third that in the second tissue sample indicates the presence of cancerous cells in the first tissue sample; methods of treating a cancerous cell by increasing the level of expression of a candidate tumor suppression gene in the cell; novel candidate tumor suppressor genes; and their use in diagnosis and therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

CANCER DIAGNOSIS AND THERAPY

Background of the Invention

This invention relates to diagnosis and treatment of cancers, particularly, solid tumors.

Sager, 246 *Science* 1406, 1989, discusses tumor suppressor genes. The loss of tumor suppressor genes, or their inactivation, is oncogenic. That is, the loss of DNA encoding a tumor suppressor gene product, or the lowering of expression of a tumor suppressor gene, gives rise to a cancerous condition. Sager generally describes the identification of candidate tumor suppressor genes. In particular, Sager describes the process of subtractive hybridization as a general method for recovering genes that are expressed in normal cells but not in closely related tumor cells. Sager further describes the isolation of three clones by subtractive hybridization of normal and cancerous mammary cells. The genes corresponding to these clones are expressed by all normal mammary epithelial cells, but not by any primary mammary tumors or mammary tumor cell lines. One such gene encodes keratin 5, which is said to be a valuable marker to distinguish normal and primary tumor cells in culture. Also identified is a gene encoding fibronectin, and a third gene identified as NB-1. Tumor suppressor genes are proposed to play a key role in cancer protection, and it is suggested that tumor suppressor genes provide a vast untapped resource for anti-cancer therapy.

Decreased DNA methylation is a consistent feature of tumorigenesis (Jones et al., *Adv. Cancer Res.* 54:1-23, 1990) but local sites of hypermethylation have also been found in tumor cells (Jones et al., *Adv. Cancer Res.* 54:1-23, 1990; Baylin et al., *Blood* 70:412-417, 1987). Elevated expression of the DNA methyltransferase gene has recently been described in progressive stages of colon

cancer (El-Deiry et al., *Proc. Natl. Acad. Sci. USA* 88:3470-3474, 1991), suggesting a general mechanism for hypermethylation, but not explaining the specificity seen on particular genes.

Summary of the Invention

This invention features novel methods for identifying cancerous cells present in a human, particularly in solid tumors. The invention also features methods for identifying drugs useful for treatment of such cancer cells, and for treatment of the cancerous condition. Unlike prior methods, the invention provides a means for identifying cancer cells at an early stage of development, such that premalignant cells can be identified prior to their spreading throughout the human body. This allows early detection of potentially cancerous conditions, and treatment of those cancerous conditions prior to spread of the cancerous cells throughout the body, or prior to development of an irreversible cancerous condition.

Tumor suppressor genes have been divided into two general types, termed class I and class II. Class I tumor suppressor genes are said to be those in which a genetic alteration (e.g., the deletion, addition or substitution of one or more nucleotides) in the coding sequence of the gene has been found to contribute to tumor cell development. In contrast, class II tumor suppressor genes are identified as those which have a lower level of expression in a cancer or precancer cell compared to a normal cell, which decreased level of expression is due to alteration in the regulation of expression of that gene, rather than to the loss of genetic information in the coding sequence of the gene. A diagnostic test based upon levels of expression of either a class I or a class II gene, or of another marker gene

that is identified as a candidate tumor suppressor gene by one of the differential screening methods described below, but which does not turn out to have tumor suppressor activity, is useful for detecting the presence of cancerous or pre-cancerous cells in a tissue sample from a patient. In addition, a patient with a cancer characterized by a lower-than-normal level of expression of one or more tumor suppressor genes can be treated (e.g., with a drug or radiation, or by transforming one or more of the tumor suppressor genes into the cancerous cells) to induce a higher level of expression of such gene(s) in the cancerous cells, thus halting or reversing the growth of the cancer.

Thus, in a first aspect the invention features a method for identifying a cancer cell in a human by providing nucleic acid from a candidate tumor suppressor gene which specifically hybridizes to RNA expressed from such a gene in a cancer cell at a level less than one third the level of hybridization with the equivalent RNA expressed from that gene in a normal cell. Alternatively, the method involves providing an antibody to the gene product of such a candidate tumor suppressor gene, which antibody specifically reacts (in the sense of an antibody-antigen reaction to form an immune complex) with the polypeptide expressed from the candidate tumor suppressor gene in a cancer cell, at a level less than one-third the level of reaction (i.e., binding) with the equivalent gene product expressed from that gene in a normal cell. The method further features obtaining from the human a tissue sample which potentially includes the cancer cell to be detected, and contacting this sample with either (1) the nucleic acid probe, under conditions which would permit hybridization with the mRNA transcribed from the gene, or (2) the antibody, under conditions appropriate for immune complex formation

between the antibody and its antigen. Finally, the method involves determining the amount of hybridization of the nucleic acid or the amount of binding of the antibody with the tissue sample, compared to the amount of hybridization of that nucleic acid or binding of that antibody with a normal tissue sample which includes only normal cells. An amount of hybridization or immune complex formation with the tissue sample less than one third the amount of hybridization or immune complex formation with the normal tissue sample is indicative of the presence of cancerous or pre-cancerous cells in the tissue sample.

The method of using a nucleic acid probe to determine the presence of cancerous cells in a tissue from a patient includes the steps of:

providing a nucleic acid probe (i.e., a single-stranded nucleic acid such as DNA, or a double stranded nucleic acid which is made single-stranded prior to doing the hybridization step) comprising a nucleotide sequence at least 8 nucleotides in length (preferably at least 15 nucleotides, and more preferably at least 40 nucleotides, and up to all or nearly all of the coding sequence) which is identical to a portion of either strand of the coding sequence of a candidate tumor suppressor gene;

obtaining from a patient a first tissue sample potentially comprising cancerous cells;

providing a second tissue sample containing cells substantially all of which are non-cancerous;

contacting the nucleic acid probe under high-stringency hybridizing conditions with RNA of each of said first and second tissue samples (e.g., in a northern blot or *in situ* hybridization assay); and

comparing (a) the amount of hybridization of the probe with RNA of the first tissue sample, with (b) the amount of hybridization of the probe with RNA of the

second tissue sample, wherein an amount of hybridization with the RNA of the first tissue sample less than one-third the amount of hybridization with the RNA of the second tissue sample indicates the presence of cancerous cells in the first tissue sample.

Alternatively, the diagnostic assay may be carried out with antibodies to the candidate tumor suppressor gene product, instead of a nucleic acid probe. Such an assay would include the following steps:

providing an antibody specific for the gene product of a candidate tumor suppressor gene, the gene product being present in cancerous tissue of a given tissue type (e.g., mammary, ovarian, bladder or prostate epithelium) at a level less than one third the level of the gene product in noncancerous tissue of the same tissue type;

obtaining from a patient a first sample of tissue of the given tissue type, which sample potentially includes cancerous cells;

providing a second sample of tissue of the same tissue type (which may be from the same patient or from a normal control, e.g. another individual or cultured cells), this second sample containing normal cells and essentially no cancerous cells; contacting the antibody with protein (which may be partially purified, in lysed but unfractionated cells, or *in situ*) of the first and second samples under conditions permitting immunocomplex formation between the antibody and any tumor suppressor gene product present in the samples; and

comparing (a) the amount of immunocomplex formation in the first sample, with (b) the amount of immunocomplex formation in the second sample, wherein an amount of immunocomplex formation in the first sample less than one third (preferably less than one fourth, and more preferably less than one tenth) the amount of

immunocomplex formation in the second sample indicates the presence of cancerous cells in the first sample of tissue.

In still another variation on the diagnostic assay of the invention, the level of a candidate tumor suppressor gene product in a biological fluid (e.g., blood or urine) of a person may be determined as a way of monitoring the level of expression of the gene in cells of that person. Such a method would include the steps of obtaining a sample of a biological fluid from the person, contacting the sample (or proteins from the sample) with an antibody specific for a candidate tumor suppressor gene product, and determining the amount of immune complex formation by the antibody, with the amount of immune complex formation being indicative of the level of the gene product in the sample. This determination is particularly instructive when compared to the amount of immune complex formation by the same antibody in a control sample taken from a normal individual or cancer patient, or in one or more samples previously or subsequently obtained from the same person.

By a candidate tumor suppressor gene is meant those genes which are found to be expressed to a significantly higher degree in normal cells than in cancerous or precancerous cells, as generally discussed above. Such a candidate tumor suppressor gene is generally identified by northern analysis or its equivalent (for example, by <u>in situ</u> hybridization) as a gene whose expression is lower in a cancer cell compared to a normal cell. If the gene bears a disabling mutation in its coding sequence, then it is termed a "candidate <u>class I</u> tumor suppressor gene". If the coding sequence of the gene is intact, inasmuch as the DNA forming the exons of that gene is not significantly altered, a southern analysis of such a gene in a cancer cell does

not reveal any significant difference in the tumor suppressor coding sequence in a cancer cell compared to a normal cell. In such genes, termed "candidate class II tumor suppressor genes", it is the regulatory mechanism of the gene that is altered in a cancerous cell compared to a normal cell.

Once the candidate class I or class II tumor suppressor gene is demonstrated to play a role in suppressing formation of tumors *in vivo* or transformation of cells *in vitro*, it may be referred to as a bona fide "class I tumor suppressor gene" or "class II tumor suppressor gene", rather than as a "candidate". Such class II genes are useful in certain of the treatment methods of the invention, because they retain a viable coding sequence which can potentially be switched on by the appropriate treatment, and such switching on will result in increased tumor suppressing activity within the treated cell. Both class I and class II genes can be transformed into cancer cells or pre-cancerous cells in order to increase their level of expression in such cells, and thus slow or prevent neoplastic growth.

By "hybridizing conditions" is meant conditions under which the nucleic acid used as a probe in the method is able to specifically hybridize with RNA expressed from a candidate tumor suppressor gene without significantly hybridizing to any other RNA expressed from either normal or cancerous human cells (e.g., conditions of high stringency, as described, for example, in Sambrook et al., Molecular Cloning, a laboratory manual, 2nd Ed., Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). In this way, hybridization of the RNA specifically indicates the presence or absence of a candidate tumor suppressor gene transcript (usually mRNA). Similarly, reaction of the antibody with the candidate tumor suppressor gene product (protein) is

performed under normal antibody-antigen reaction conditions which allow specific recognition of the candidate tumor suppressor gene product by antibody, with little or no cross-reaction of the antibody with other proteins normally present in the cancerous or normal cells. In this way, measurement of the amount of antigen-antibody immune complex formed in the sample is indicative of the amount of candidate tumor suppressor gene product present in that sample.

In preferred embodiments, the candidate tumor suppressor gene is a gene encoding keratin 5, NB-1 gene product, fibronectin, connexin 26, glutathione-S-transferase pi, CaN19 protein (formerly called clone 19 gene product), small proline-rich (spr-1) protein, amphiregulin, thymosin beta-4, gamma actin, calpactin light chain (p11), HBp17, myosin regulatory light chain, V-Fos transformation effector protein, or one of the following mitochondrial genome-encoded proteins: URF4, Co III, and ATPase6. The candidate tumor suppressor gene may alternatively be one of the newly-identified genes herein referred to as U1-U10, partial sequences of which are given as SEQ ID NOs: 3-12. These genes are specifically described in detail below.

In other preferred embodiments, the amount of hybridization of the nucleic acid with RNA from precancerous or cancerous cells in the human is less than one third the level detected with a normal cell, more preferably less than one tenth the level, or even more preferably is undetectable.

In a second aspect, the invention features a method for identifying a drug useful for treatment of a cancer cell. The method includes the steps of

identifying a candidate class II tumor suppressor gene, expression of which is suppressed [i.e., significantly diminished (e.g., by two thirds or more)]

in a given type of cancerous cell from a given type of tissue, compared to a normal cell in the same type of tissue;

providing a first and a second sample of that given type of cancerous cell;

treating the second sample with a candidate drug; and

determining the level of expression of the gene in the second sample after treatment with the candidate drug, wherein a drug which increases the level of expression of the gene in the second sample, compared to the level of expression of the gene in the untreated first sample, is potentially useful for treatment of the given type of cancer cell, and perhaps for other types of cancer cells, as well.

Generally, the candidate class II tumor suppressor gene and the level of expression of that gene are identified or determined as discussed herein. Potentially useful drugs may be chosen from, for example, those which alter signal transduction pathways, or which facilitate demethylation of methylated residues on DNA. Such drugs may increase tumor suppressor gene expression by, for example, increasing tumor suppressor gene messenger RNA synthesis, or mRNA processing, or protein synthesis, or by decreasing RNA degradation or protein degradation.

In a related aspect, the invention features methods for treating a patient who has cancer. One such method involves the steps of identifying, in a human, a cell having a low level of expression of a candidate tumor suppressor gene compared to a normal cell, and treating that cell with a drug identified as one which raises the level of expression of that candidate tumor suppressor gene in the cell. Stated another way, the method of treatment includes the steps of identifying a

patient with a cancer cell characterized by a low level of expression of a candidate tumor suppressor gene, compared to the level of expression of such gene in a normal cell of the same tissue type as the cancer cell; and either treating the cancer cell with a compound which raises the level of expression of the gene in the cancer cell, or introducing into the cancer cell a nucleic acid encoding the gene. Preferably, the nucleic acid would include an expression control element permitting expression of the gene in the cancer cell. Treating patients with such drugs or gene therapy provides a means to control or eliminate their cancers.

The invention also includes an isolated DNA which hybridizes under high-stringency conditions to any one of the sequences shown as SEQ ID NOs: 3-12, including but not limited to an isolated DNA which has a sequence identical to any one of SEQ ID NOs: 3-12. The term "isolated DNA" denotes a DNA that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the candidate tumor suppressor gene that hybridizes to the sequence shown in the applicable SEQ ID NO. The term therefore includes, for example, a cDNA encoding the applicable candidate tumor suppressor gene product; a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote; or a genomic DNA fragment produced by PCR or restriction endonuclease treatment. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Detailed Description

The drawing is first briefly described.

Drawing

The Figure illustrates the DNA sequence of a cDNA encoding human connexin 26 (Cx26), and the amino acid sequence deduced therefrom.

Candidate Tumor Suppressor Genes

Candidate class I and class II tumor suppressor genes are generally described above. These candidate tumor suppressor genes can be identified as described by Sager, supra, or as described by Trask et al., 87 *Proc. Natl. Acad. Sci. USA* 2319, 1990; Yaswen et al., 87 *Proc. Natl. Acad. Sci. USA* 7360, 1990; and Lee et al., 88 *Proc. Natl. Acad. Sci. USA* 2825, 1991, wherein specific subtractive hybridization methods are provided. All of these publications are herein incorporated by reference. In addition, the subtractive hybridization method described below may be used. The subtractive hybridization method is particularly advantageous in screening for candidate tumor suppressor genes since it provides a positive selection procedure.

The following is a specific example of such a subtractive hybridization procedure used to screen for candidate tumor suppressor genes involved in breast cancer. This example is not limiting in the invention and those of ordinary skill in the art will recognize that many variations to this method can be used with equivalent efficacy in identifying useful candidate tumor suppressor genes.

Identification and Isolation of Candidate Genes

In this example the medium, DFCI-1, described by Band and Sager, 86 *Proc. Natl. Acad. Sci., USA* 1249, 1989 was used because of its ability to support similar growth

of both normal and tumor-derived human mammary epithelial cells. cDNA rather than genomic DNA was used for screening since the cDNAs are smaller and easier to manipulate than their genomic counterparts, and are present in multiple copies. Recovery of such cDNAs allows their use as probes to isolate the equivalent genomic DNA. Further, the cDNA can be expressed in a expression vector to produce the tumor suppressor gene product, and thus allow production of antibodies to that product for use in the methods described herein. The cDNA may be genetically manipulated so that it encodes only a chosen portion of the full-length gene product, resulting in expression of a defined oligopeptide fragment of the tumor suppressor gene product that may be used to generate antibodies useful for detecting the full-length gene product. Alternatively, the oligopeptides may be chemically synthesized. Design and production of such defined fragments may be accomplished by standard methods.

The normal cells used in the methods described herein were derived from a strain 76N established from discarded reduction mammoplasty tissue as described by Band and Sager, supra. These cells are diploid and senesce after 15-20 passages. The tumor cells were derived from an aneuploid cell line established from a pleural effusion as described by Band et al., 1 *Genes, Chromosomes, and Cancer* 48, 1989 and Band et al., 50 *Cancer Research* 7351, 1990. However, any cells used for subtractive hybridization can be derived from any individuals, and substituted as described below. Primary tumor cells or metastatic cells can be used. In this example, both parental cell populations were grown in DFCI-1 medium at similar population doubling times of about 30 hours. These cells were harvested at 70% confluency directly into 4M guanidium isothiocyanate,

0.5M sodium citrate, and 0.1M β-mercaptoethanol for RNA preparation. Total RNA was extracted from the cells by lysis in the guanidium isothiocyanate mixture, and poly(A)+RNA purified by two cycles of affinity chromatography on oligo(dT) cellulose by standard technique. The cDNA was synthesized using Moloney murine leukemia virus reverse transcriptase from Bethesda Research Laboratories with an oligodeoxynucleotide oligo$(dT)_{12-18}$ as a primer.

The ^{32}P pre-labeled SS cDNA from 76N cells was hybridized with a 10-fold excess of tumor poly$(A)^+$ mRNA from 21MT-2 cells (Band et al., *Cancer Research* 1990). 500 ng fibronectin (FN) mRNA, prepared by in vitro transcription was added to subtract out FN cDNA, which is present at high abundance in the mRNA of the normal cells. The hybridization reaction mixture was loaded onto a hydroxylapatite column maintained at 60°C and eluted with 0.1M phosphate buffer (pH 6.8). After rerunning the effluent through the column three times, the effluent was collected and rehybridized as above (2nd subtraction) without added FN mRNA. The final effluent was concentrated to 100 µl, a sample was removed for quantitation, and the rest frozen for subsequent screening.

cDNA from 76N poly$(A)^+$ RNA was used to produce a recombinant library in the phagemid lambda Zap II (Stratagene Corp., La Jolla, CA) by procedures recommended by the vender. The 76N library was screened by differential hybridization using the ^{32}P random-primer labelled subtracted cDNA probe against the tumor specific cDNA. After a secondary screening the differentially expressed clones were isolated, and the inserts were amplified by PCR from phage using T3 and T7 sequences as primers. After gel electrophoresis, the PCR products were purified by phenol/chloroform extraction from

agarose and ^{32}P random-primer labelled for RNA northern analysis.

Total RNA (20ug) was heat denatured at 68°C for 15-20 min. followed by electrophoresis in 1.2% agarose-formaldehyde gels and transferred to nylon membranes (Zeta-probe, BioRad); prehybridization and hybridization were performed as described by Haskill et al., 87 *Proc. Natl. Acad. Sci, USA* 7732, 1990. Sequencing of cloned DNA was performed either directly or on exonuclease III-deleted derivatives. These deletion derivatives were generated using a Promega Erase-a-Base kit but can be generated by using other standard technique. Sequencing was carried out by a dideoxy chain termination method with T7 DNA polymerase (Pharmacia). Parallel reactions were also performed with dGTP analogs (Pharmacia) when necessary to resolve sequence compressions.

In one subtraction, 50 clones were recovered. After two rounds of screening, seven different clones showed unique or highly preferential expression in normal cells compared to tumor cells. The clones were identified by northern hybridization using standard techniques. The size range of mRNAs varied from 0.6 kb to almost 5 kb. These clones include genes expressed at rare to high abundance in mRNAs.

One clone, termed clone 1-3, is expressed in four normal strains but not in a series of tumor-derived lines. It has been shown by sequence comparison in GENBANK to encode the human homolog of rat connexin 26 (Cx26), a gap junction protein the DNA sequence of which is provided by Zhang and Nicholson 109 *Journal Cell Biology* 3391, 1989. The DNA and deduced amino acid sequence of human connexin 26 (SEQ ID NO: 2) is given below in the Figure. This cDNA clone has a single long open reading frame that extends to a stop codon at base 881, and encodes a putative protein of 226 amino acid

residues with a predicted molecular mass of ~26,000 daltons.

Preceding the initiator ATC, 23 nucleotides upstream from ATG, is a consensus splice accepter signal (TTTCCAG), raising the possibility that splicing occurs at this site to create two sizes of human Cx26 transcripts. This signal sequence is not present in the 5' region of the rat Cx26 sequence, which does not produce two transcripts. The 3' untranslated regions contains a possible polyadenylation signal sequence AATAAA positioned 87 nucleotides upstream from the poly(A)$^+$ tail. At nucleotide positions 1326, 1623, 1664 and 2082 a putative instability sequence ATTTA, is present, which may be involved in posttranscriptional regulation. The overall nucleotide homology between human and rat Cx26 is 86.2% within the open reading frame. The amino acid sequence deduced from the human cDNA is 92.5% identical to rat Cx26. However, the 5' and 3' untranslated regions show no significant similarity between human and rat.

To confirm the intracellular location of gap junction proteins in human mammary epithelial cells, we examined cells by immunofluorescence using anti-Cx26 or anti-Cx43 antibodies. Specific fluorescent spots were found at membrane contact sites of 76N cells (a normal human mammary epithelial cell line), whereas no fluorescence staining was observed with 21MT2 cells (a human breast tumor cell line). When fixed cells were treated with preimmune serum, the immunoreactivity failed to show discrete punctate staining at the cell membrane. Failure to localize Cx26 or Cx43 protein at the junctional areas of 21MT2 cells is consistent with the lack of connexin mRNA expression observed in breast tumor cells.

To assess the relative periodicity of connexin gene expression during the cell cycle, normal mammary epithelial cells were synchronized in G_1 by lovastatin (15 μM/24 hours), released from lovastatin-induced arrest by the addition of 2 mM mevalonate, and then sampled at 3 hour intervals over the next 33 hours. Cx26 and Cx43 transcript levels were analyzed by Northern blot analysis of total RNA prepared from samples taken at indicated times. The progress of the cells through the cell cycle was monitored by [^{3}H]thymidine incorporation and by the level of histone H4 mRNA in Northern blot analysis. Histone H4 was induced in S phase at 18 hour. The time of appearance of histone H4 message coincided with the peak period of DNA synthesis as measured by [^{3}H]-thymidine incorporation. The upper Cx26 transcript increased at 6 hr. to a moderate steady state level until 21 hr, near the end of S phase, when both Cx26 transcripts showed a further increase in G_2. In contrast to Cx26 mRNA, Cx43 epxression during the cell cycle was relatively invariant. Considering the assumed similarity of their functions, the expression of both connexins during the cell cycle might be expected to show a similar regulation pattern. Thus, the difference in cell cycle regulation of Cx26 and Cx43 is quite surprising.

Connexins are structural proteins that surround the channels of which gap junctions are composed; the channels in turn provide direct communication between adjacent cells. Gap junctions have been postulated to play a growth regulatory role, on the basis of numerous correlations between growth control and junctional communication. Of these, one of the earliest and still the most striking is Stoker's experiment in which polyoma-transformed BHK cells were inhibited from colony formation by contact (later shown to be junctional communication) with a monolayer of normal BHK cells.

Recent experiments by Loewenstein and coworkers and others have correlated post-translational modulation of junctional communication with growth inhibition. Our results, in contrast, suggest transcriptional regulation. This opens the possibility for experimental and clinical modulation at the level of transcription as described below.

A clone termed clone 2-3 encodes glutathione-S-transferase pi, identified by sequence comparison with known genes in GENBANK. The DNA sequence for glutathione-S-transferase pi is provided by Moscow et al., 49 *Cancer Research* 1422, 1989. This protein is a well-characterized enzyme, present in many cell types, that has detoxifying activity against many lipophilic toxic agents including carcinogens. We have found that it is down-regulated in a number of mammary tumor-derived cell lines, both primary and metastatic, but strongly expressed in normal and immortalized mammary epithelial cells grown in culture.

A clone originally termed clone 19, and now referred to as CaN19, represents a gene expressed in normal mammary epithelial cell strains but not in tumor-derived cell lines. The DNA sequence (and corresponding amino acid sequence, or "gene product") of CaN19 is shown as SEQ ID NO: 1 below. Sequence comparisons have shown that CaN19 is a member of the S100 gene family, encoding small Ca^{++} binding proteins (about 10 kD) with diverse functions. These proteins have two "EF hands", domains where Ca^{2+} is bound, in contrast to calmodulin proteins which have four. The S100 beta protein is a major constituent of glial cells, whereas related proteins are expressed in differentiated but not in undifferentiated PC 12 (rat pheochromocytoma) cells. CaN19 is also related in structure to the small regulatory subunit of calpactin, p11. MRP8 and MRP14 are also related and are

S100 proteins expressed by macrophages during chronic inflammation. Calabretta et al., 261 *J. Biol. Chem.* 12628, 1986. Another related protein, calcyclin, has been found in serum-induced cycling cells, but not in quiescent cells, and in leukocytes from CML patients. A related mouse protein is also cell cycle induced. The possibility that calcyclin expression might be cancer related is particularly interesting in view of our evidence that CaN19 is not expressed in breast tumor cells. CaN19 appears to be negatively regulated in tumors, in contrast to calcyclin. Other related proteins are described by Kligman and Hilt 13 *TIBS* 437, 1988.

Other genes which are useful in the present invention include NB-1 described by Yaswen et al., supra; keratin 5 as described by Trask et al., supra, the DNA sequence of which is published in 8 *Molecular Cell Biology* 486, 1988; and small proline-rich protein (spr-1), the sequence of which is published in 18 *Nucl. Acid Res.* 4401-4407, 1990. The latter gene is known to be expressed at higher levels following treatment with ultraviolet radiation, suggesting that the protein may have a DNA repair function. Thus, spr-1 is a very promising gene for further investigation.

In further experiments, an adaptation of the subtractive hybridization technique was used which proved to be less laborious and more efficient for cloning of candidate tumor suppressor genes, including rarely expressed genes, than the hydroxyapatite column method. This method utilizes a biotinylation-based subtraction procedure (Schweinfest et al., 7 *Genet. Annal. Techn. Appl.* 64-70, 1990; Swaroop et al., 19 *Nucl. Acids Res.* 1954, 1991), instead of hydroxyapatite as previously used. In this procedure, a single strand phagemid cDNA library from normal cell polyA$^+$ mRNA is hybridized with excess biotinylated tumor polyA$^+$ mRNA, and the resulting

double stranded sequences are removed by binding to streptavidin. The remaining single-stranded phagemid cDNAs are converted to double-stranded form and used to transform bacterial host cells. The resulting subtracted cDNA library is differentially screened with total cDNA from normal and tumor cells. This method produced some 20 additional cloned cDNAs, including some which, upon partial sequencing, proved to have been previously identified by others, and some which appear to be novel. The previously-identified genes which were found by this method to be candidate tumor suppressor genes potentially useful in the methods of the invention include genes encoding human amphiregulin (the full sequence of which can be found in GENBANK at locus HUMARXC, Accession #M30704); thymosin beta-4 (locus HUMTHYB4, Accession #M17733); gamma actin (locus HUMACTCGR, Accession ##X04098, K00791, M24241); calpactin light chain (p11) (locus HUMCALPA1L, Accession #M81457); HBp17 (locus HUMHEPBP, Accession #M60047), myosin regulatory light chain (locus HUMMRLCM, Accession #X54304); v-fos transformation effector protein (locus HUMFTE1A, Accession #M84711); and the mitochondrial genome-encoded proteins URF4 (locus HUMMTHSXX, Accession #V00662); Co III (locus HUMMTHSXX, Accession #V00662); and ATPase6 (locus HUMMTCG, Accession ##J01415, M12548, M58503, M63932, and M639333). Also found by this method were several genes which, on the basis of the partial DNA sequences set forth as SEQ ID NOs: 3-12, respectively, appear to be novel sequences not previously entered into GENBANK. The portion of the cDNAs so sequenced represents part of the coding region and/or part of the 3' untranslated region of each cDNA. Still other genes can be identified as described above using northern analysis of isolated clones and determining whether tumor

cell expression of the gene is reduced by at least 2/3 compared to normal cells.

Most or all of the genes described herein have been found to be expressed at a low but detectible level in at least some tumor cells (which is taken to be an indication that the coding sequence is intact in these cells), and thus appear to be candidate class II rather than class I tumor suppressor genes in these tumors. Another indication that a particular candidate tumor suppressor gene falls within class II in a particular tumor is a normal-appearing Southern blot of the tumor's genomic DNA when probed with the tumor suppressor gene cDNA. (Small deletions or rearrangements might not be detected, of course.) Candidate class I tumor suppressor genes, in which the coding sequence of the gene is altered in a way to yield no biologically active gene product or an altered gene product, could also be detected by the differential hybridization screening method of the invention if the genetic alterations are such that (1) no detectable mRNA is transcribed from the mutant gene, or (2) the mRNA transcribed from the gene is sufficiently different from wild type that it cannot hybridize to the hybridization probe utilized, or (3) the mRNA has an altered sequence resulting in a different location on a Northern gel than the normal mRNA, or (4) the mRNA is hydrolyzed by the cell rapidly after transcription. The alternative method of detection disclosed herein, in which an antibody to the wild type candidate tumor suppressor gene product is used to detect gene expression in cell samples, would also be useful for identifying candidate Class I tumor suppressor genes, and for detecting their expression in a given cell sample, if the mutations in the coding sequence of the gene are such that (1) no stable gene product is expressed by the

mutant gene, or (2) the gene product that is expressed is so altered that the antibody utilized cannot bind to it.

Diagnostic applications

Class II genes are of particular interest because the suppressor gene has not been lost, and may therefore be available for up-regulation by drugs or other treatment. Restoration of suppressor gene function by regulatory intervention offers new opportunities in the design of novel drugs for cancer therapy.

Both Class I and Class II genes are immediately valuable for early diagnosis and prognosis, which are especially pressing needs in breast cancer where the course of the disease is so unpredictable. Some genes expressed preferentially in normal cells may not have tumor suppressor functions. They are nonetheless useful as diagnostic markers.

The candidate suppressor genes described herein represent just the "tip of the iceberg" with respect to loss-of-function genes that may be useful in diagnosis, prognosis, and therapy. Genes with numerous and diverse functions are anticipated to participate in protecting the long-lived human species from cancer. They include DNA repair genes that maintain genomic integrity and stability, genes that promote irreversible steps in differentiation, and genes that regulate proliferation. Cancer starts at the cellular level, but becomes a systemic disease, and at that point, systemic mechanisms of protection play important roles. These include cell-cell communication by gap junctions, paracrine regulation by growth factors and cytokines, protection by the immune system, control of angiogenesis, and the regulation of tumor invasion. For each of these, specific genes encode key proteins whose loss may facilitate neoplasia. The experimental system described herein allows early

recognition of aberrant tumor suppressor and diagnostic genes.

As discussed above, both candidate class I and candidate class II tumor suppressor genes can be used for diagnosis of cancer. All of those genes described above, and other genes identified in a similar manner, are potentially useful for diagnosis of cancerous conditions. For example, they are particularly useful for identification of cancerous cells in solid tumors, such as in breast cancer. Once a lump is detected in a mammogram, or by other means in a breast, a portion of that lump may be removed and analyzed by northern analysis or by in situ hybridization using the cloned gene (or antibodies to the gene product produced by standard techniques) to determine whether the level of expression of the candidate tumor suppressor gene is normal or at a reduced level. If it is at a reduced level, this will be indicative that the cells in that lump are cancerous or pre-cancerous and appropriate steps may be taken to either remove or treat those cells in vivo.

Similarly, routine diagnosis can be obtained in a manner similar to a papsmear in which cells are taken from a human and tested by hybridization with any one or more of the above candidate tumor suppressor genes or by immune complex formation with antibodies to the gene products. Such testing will allow earlier diagnosis of cancerous conditions than has previously been possible.

Those of ordinary skill in this art will recognize that the northern analysis and in situ hybridization or immune complex formation can be carried out by any of a number of standard techniques. For example, the DNA of a candidate tumor suppressor gene or its equivalent cDNA may be used as a probe for RNA transcribed from those genes in cells to be tested. Similarly, DNA which

hybridizes to the RNA produced by such genes can also be used.

The cDNA or its equivalent may be placed in expression vectors to cause production of candidate tumor suppressor gene products which may be purified and used to isolate polyclonal or monoclonal antibodies to those candidate tumor suppressor gene products. Those particular antibodies which are specific for (i.e., form readily detectible immune complexes with) the candidate tumor suppressor gene product can be identified by standard procedures. Generally, it is preferred that a specific monoclonal antibody be identified so that a large amount of that antibody can be readily produced and used in diagnostic procedures. Immunoprecipitation by antibodies of candidate tumor suppressor gene products is performed by standard methodology such as western blotting.

These diagnostic methods can be adapted for use as a way to monitor changes in the level of expression of a given candidate tumor suppressor gene in a given patient over time. This would be useful, for example, as a routine measure for monitoring for the presence of cancer in apparently healthy subjects, much as pap smears and mammograms are used. This technique relies upon the normal expression of a given candidate tumor suppressor gene product in a readily obtainable biological fluid such as blood, urine, or saliva. A baseline normal level of expression of the gene product would be established by analyzing samples taken from the subject over the years, or by comparison with standards obtained from other, disease-free individuals. A drop in the amount of the gene product present in a given sample would be an indication of the presence of tumor cells in the subject. Alternatively, the method could be adapted to serve as a means for following the clinical progression of a tumor,

wherein increases or decreases in the level of the gene product in the analyzed sample would be indicative of decreasing or increasing tumor load, respectively.

The above-described method for assaying a biological fluid will work reliably only if the candidate tumor suppressor gene product is normally a secreted protein. Whether or not a given gene product is secreted can be determined empirically (e.g., by using an antibody specific for the gene product), or may be predicted by the presence of a secretion signal sequence in the cDNA (e.g., as taught by Von Heijne, 133 *Eur. J. Biochem.* 17-21, 1983) in accordance with standard methods.

Screening for and Treatment with Transcription-activating Drugs

As generally discussed above, candidate class II tumor suppressor genes can be used to identify useful drugs for treatment of cancers. This may be accomplished by standard procedures by culturing cells which include tumor suppressor genes (which are either expressed at normal or subnormal levels) and treating those cells with a variety of drugs to determine which drugs increase the level of expression of the candidate tumor suppressor gene product within those cells. It is preferred that a cancerous cell be used in such a procedure since the increased level of expression of the candidate tumor suppressor gene product will be more readily detected in such a cell, and the drug may work only on genes the expression of which is lower than normal. Identification of the increase in tumor suppressor gene expression can be analyzed by standard northern or in situ analysis or by antibody testing. Alternatively, rather than looking for expression of the tumor suppressor gene, the concomitant increase in a function of that gene may be detected by standard techniques. Two examples

illustrating such a procedure are given below. In these examples, phorbol myristate acetate (PMA) is found to increase expression of the Cx26 tumor suppressor gene in tumor cells but not in normal cells, while azadeoxycytidine increases the level of expression of the CaN19 candidate tumor suppressor gene in tumor cells but not in normal cells.

Once the appropriate drug is identified, it may be administered to humans who are identified as containing cells having a reduced level of the tumor suppressor gene product. This may be accomplished either by direct administration of the drug at the tumor site or by systemic treatment with the drug.

Drug-induced Stimulation of CaN19 mRNA Expression

In an analysis of the effects of certain drugs on induction of CaN19 expression, exponentially growing normal and tumor cells were treated with the following agents (5Br-cAMP, 1mM; forskolin, 10uM; PMA, 100ng/ml; retinoic acid, 1uM; A23187, 0.5uM; actinomycin D, 5ug/ml; cycloheximide, 10 ug/ml; okadaic acid, 5ng/ml; TGF-β, 1ng/ml; prolactin, 1mg/ml; β-estradiol, 2nM; 5-aza-2'-deoxycytidine, 1uM-100uM; all purchased from Sigma Chemical Co. except TGF-β from Collaborative Research Inc.) (Lee et al., *Mol. Cell. Biol.* 10:1982-1988, 1990). To study the effect of azadeoxycytidine, cells were plated at low density (~25% confluency) and incubated in the presence of various concentrations of drug. Cells were washed, retreated with drug in fresh medium for another 2 days, and then harvested for RNA analysis (~70% confluency). The steady state levels of mRNA were examined by Northern blot analysis with RNA extracted from normal and tumor cells at different time points (0, 1, 3, 6, 12, and 24 hours) after each drug treatment. Additions of 5Br-cAMP, forskolin, PMA, retinoic acid,

actinomycin D, cycloheximide, A23187, or okadaic acid were without noticeable effect on the level of expression of CaN19 mRNA in tumor cells. In contrast, exposure of mammary tumor cells to azadeoxycytidine induced the expression of CaN19-specific RNA. The level of expression of CaN19 in normal cells was not affected by azadeoxycytidine treatment. These findings suggest that DNA methylation plays a direct role in control of CaN19 gene expression in tumor cells. Since aza-dCyd is a well-established DNA demethylating agent, it is very likely that treatment with this drug demethylated transcription binding sites in CaN19 and possibly in other unidentified genes as well. Although systemic treatment with aza-dCyd itself is said to be toxic and tumorigenic (Harrison et al., *Proc. Natl. Acad. Sci. USA* 80:6606-6610, 1983; Carr et al., *Carcinogenesis* 5:1583-1590, 1984), these results provide insight into a possible mechanism for switching on candidate tumor suppressor genes in tumor cells, and suggest testing other DNA demethylating agents for antitumor potential.

Drug-induced Stimulation of Cx26 mRNA Expression

Two different breast cancer cell lines, one from a primary tumor and one from a metastatic cell line, were found to have significantly reduced levels (compared to levels in normal breast cells) of connexin 26 expression by northern analysis, as discussed above. A short treatment of these cells with phorbol myristate acetate (PMA) induced the expression of mRNA in these cells, while treatment with certain other drugs that affect signal transduction pathways was found to have no effect on Cx26 expression in these cells at the concentrations tested. Specifically, growing 21 PT cells (Band et al., *Cancer Research* 1990; derived from a primary tumor) were treated with 100 ng/ml PMA, 1 mM dBc-cAMP, 1μM retinoic

acid, 5μg/ml actinomycin D, 10μg/ml cycloheximide, 5ng/ml okadaic acid, 2nM β-estradiol, or 1ng/ml TGFβ at time zero in a series of dishes. At time points from 0 to 48 hours after exposure to drug, samples were taken for RNA extraction and northern blot analysis. In the PMA-treated cells, expression of connexin 26 was observed by 3 hours, peaking at 6-12 hours (at 25% normal cell levels) and decreasing by 24 hours. Similar results were obtained with 21MT-2 cells, another tumor cell line. In contrast, PMA treatment of normal cells did not increase the level of Cx26 gene expression above control levels.

In order to see whether Cx26 mRNA stimulation in PMA-treated tumor cells leads to protein synthesis, immunofluorescence staining with anti-Cx26 antibody and scrape-loading dye transfer experiments were performed at various times after PMA treatment, using several mammary tumor cell lines. Cx26 proteins were not detected at cell to cell junctional areas nor was junctional communication detected between cells. Neither method was sensitive enough to detect a very weak signal, which might have resulted from the short half-life of the induced mRNA.

Gene Therapy

As generally discussed above, tumor suppressor genes of both class I and class II may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Although such gene therapy is particularly appropriate for use in cells, both cancerous and precancerous, in which the level of a particular tumor suppressor gene product is diminished compared to normal cells, it may also be useful to increase the level of expression of a given tumor suppressor gene even in those tumor cells in

which the gene is expressed at a "normal" but perhaps not optimal level.

21MT-2 cells, a line of cultured breast tumor cells (developed in this laboratory) in which the level of Cx26 mRNA is undetectible, were transfected with a plasmid construct containing the full-length cDNA corresponding to Cx26 linked to appropriate expression control elements. Unlike the untransfected cells, the transfectants expressed significant amounts of Cx26 protein. Furthermore, the transfected cells were found to assemble the Cx26 protein into gap junctions that functioned in cell-cell communication in the same manner as described for normal mammary epithelial cells. These results indicate that, by transferring a candidate tumor suppressor gene along with expression control elements into a tumor cell which does not express the gene from its own genome, tumor cells can be induced to produce functional candidate tumor suppressor gene product at high levels.

Gene therapy would be carried out according to generally accepted methods: for example, as described by Friedmann in *Therapy for Genetic Disease*, T. Friedman (ed.), Oxford Univ. Press, 1991, pp.105-121. Cells from a patient's tumor would first be analyzed by the diagnostic methods described above, in order to ascertain which if any of the candidate tumor suppressor genes are expressed at a significantly lower than normal level (or not at all) in the tumor cells. A virus or plasmid containing a copy of such a tumor suppressor gene linked to expression control elements and capable of replicating inside the tumor cells would then be injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of

each of the targeted tumor cells, the treatment may have to be repeated periodically.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Sager, Ruth

(ii) TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Fish & Richardson
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: U.S.A.
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
(B) COMPUTER: IBM PS/2 Model 50Z or 55SX
(C) OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
(D) SOFTWARE: WordPerfect (Version 5.0)

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 07/552,216
(B) FILING DATE: February 28, 1991

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Fraser, Janis K.
(B) REGISTRATION NUMBER: 34,819
(C) REFERENCE/DOCKET NUMBER: 00530/048002

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (617) 542-5070
(B) TELEFAX: (617) 542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	452
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	double
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GCACGAGCTG GGTCTGTCTC TGCCACCTGG TCTGCCACAG ATCCATG ATG TGC AGT      56
                                                      Met Cys Ser
                                                        1

TCT CTG GAG CAG GCG CTG GCT GTG CTG GTC ACT ACC TTC CAC AAG TAC      104
Ser Leu Glu Gln Ala Leu Ala Val Leu Val Thr Thr Phe His Lys Tyr
  5                  10                  15

TCC TGC CAA GAG GGC GAC AAG TTC AAG CTG AGT AAG GGG GAA ATG AAG      152
Ser Cys Gln Glu Gly Asp Lys Phe Lys Leu Ser Lys Gly Glu Met Lys
20                  25                  30                  35

GAA CTT CTG CAC AAG GAG CTG CCC AGC TTT GTG GGG GAG AAA GTG GAT      200
Glu Leu Leu His Lys Glu Leu Pro Ser Phe Val Gly Glu Lys Val Asp
                40                  45                  50

GAG GAG GGG CTG AAG AAG CTG ATG GGC AAC CTG GAT GAG AAC AGT GAC      248
Glu Glu Gly Leu Lys Lys Leu Met Gly Asn Leu Asp Glu Asn Ser Asp
            55                  60                  65

CAG CAG GTG GAC TTC CAG GAG TAT GCT GTT TTC CTG GCA CTC ATC ACT      296
Gln Gln Val Asp Phe Gln Glu Tyr Ala Val Phe Leu Ala Leu Ile Thr
        70                  75                  80

GTC ATG TGC AAT GAC TTC TTC CAG GGC TGC CCA GAC CGA CCC              338
Val Met Cys Asn Asp Phe Phe Gln Gly Cys Pro Asp Arg Pro
    85                  90                  95


TGAAGCAGAA CTCTTGACTC CCTGCCATGG ATCTCTTGGG CCCAGGACTG TTGATGCCTT    398

TGAGTTTTGT ATTCAATAAA CTTTTTTTGT CTGTTGAAAA AAAAAAAAAA AAAA          452
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	2261
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	double
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
GATTTAATCC TATGACAAAC TAAGTTGGTT CTGTCTTCAC CTGTTTTGGT GAGGTTGTGT      60

AAGAGTTGGT GTTTGCTCAG GAAGAGATTT AAGCATGCTT GCTTACCCAG ACTCAGAGAA     120

GTCTCCCTGT TCTGTCCTAG CTAGTGATTC CTGTGTTGTG TGCATTCGTC TTTTCCAGAG     180

CAAACCGCCC AGAGTAGAAG ATG GAT TGG GGC ACG CTG CAG ACG ATC CTG GGG     233
                      Met Asp Trp Gly Thr Leu Gln Thr Ile Leu Gly
                        1               5                  10

GGT GTG AAC AAA CAC TCC ACC AGC ATT GGA AAG ATC TGG CTC ACC GTC       281
Gly Val Asn Lys His Ser Thr Ser Ile Gly Lys Ile Trp Leu Thr Val
            15                  20                  25

CTC TTC ATT TTT CGC ATT ATG ATC CTC GTT GTG GCT GCA AAG GAG GTG       329
Leu Phe Ile Phe Arg Ile Met Ile Leu Val Val Ala Ala Lys Glu Val
        30                  35                  40

TGG GGA GAT GAG CAG GCC GAC TTT GTC TGC AAC ACC CTG CAG CCA GGC       377
Trp Gly Asp Glu Gln Ala Asp Phe Val Cys Asn Thr Leu Gln Pro Gly
    45                  50                  55

TGC AAG AAC GTG TGC TAC GAT CAC TAC TTC CCC ATC TCC CAC ATC CGG       425
Cys Lys Asn Val Cys Tyr Asp His Tyr Phe Pro Ile Ser His Ile Arg
60                  65                  70                  75

CTA TGG GCC CTG CAG CTG ATC TTC GTG TCC AGC CCA GCG CTC CTA GTG       473
Leu Trp Ala Leu Gln Leu Ile Phe Val Ser Ser Pro Ala Leu Leu Val
                80                  85                  90

GCC ATG CAC GTG GCC TAC CGG AGA CAT GAG AAG AAG AGG AAG TTC ATC       521
Ala Met His Val Ala Tyr Arg Arg His Glu Lys Lys Arg Lys Phe Ile
            95                  100                 105

AAG GGG GAG ATA AAG AGT GAA TTT AAG GAC ATC GAG GAG ATC AAA ACC       569
Lys Gly Glu Ile Lys Ser Glu Phe Lys Gln Ile Glu Glu Ile Lys Thr
        110                 115                 120

CAG AAG GTC CGC ATC GAA GGC TCC CTG TGG TGG ACC TAC ACA AGC AGC       617
Gln Lys Val Arg Ile Glu Gly Ser Leu Trp Trp Thr Tyr Thr Ser Ser
    125                 130                 135
```

```
ATC TTC TTC CGG GTC ATC TTC GAA GCC GCC TTC ATG TAC GTC TTC TAT      665
Ile Phe Phe Arg Val Ile Phe Glu Ala Ala Phe Met Tyr Val Phe Tyr
140                 145                 150                 155

GTC ATG TAC GAC GGC TTC TCC ATG CAG CGG CTG GTG AAG TGC AAC GCC      713
Val Met Tyr Asp Gly Phe Ser Met Gln Arg Leu Val Lys Cys Asn Ala
                160                 165                 170

TGG CCT TGT CCC AAC ACT GTG GAC TGC TTT GTG TCC CGG CCC ACG GAG      761
Trp Pro Cys Pro Asn Thr Val Asp Cys Phe Val Ser Arg Pro Thr Glu
            175                 180                 185

AAG ACT GTC TTC ACA GTG TTC ATG ATT GCA GTG TCT GGA ATT TGC ATC      809
Lys Thr Val Phe Thr Val Phe Met Ile Ala Val Ser Gly Ile Cys Ile
        190                 195                 200

CTG CTG AAT GTC ACT GAA TTG TGT TAT TTG CTA ATT AGA TAT TGT TCT      857
Leu Leu Asn Val Thr Glu Leu Cys Tyr Leu Leu Ile Arg Tyr Cys Ser
    205                 210                 215

GGG AAG TCA AAA AAG CCA GTT                                         878
Gly Lys Ser Lys Lys Pro Val
220                 225
```

```
TAAATTGCCC AGTTGTTAGA TTAAGAAATA GACAGCATGA GAGGGATGAG GCAACCCGTG    938

CTCAGCTGTC AAGGCTCAGT CGCCAGCTCC CAACACAAAG ATTCTGACCT TAAATGCAAC    998

CATTTGAAAC CCCTGTAGGC CTCAGGTGAA ACTCCAGATG CCACAATGAG CGCTCCCCTA   1058

AAGCCTCAAA ACAAAGGCCT AATTCTATGC CTGTCTTAAT TTTCTTTCAC TTAAGTTAGT   1118

TCCACTGAGA CCCCATGTTA GGGGTTATTG GTGTAAGGTA CTTTCATATT TTAAACAGAG   1178

GATATCGGCA TTTGTTTCTT TCTCTGAGGA CAAGAGAAAG CCAGGTTCCA CAGAGGACAC   1238

AGAGAAGGTT TGGGTGTCCT CCTGGGGTTC TTTTTGCCAA CTTTCCCCAC GTTAAAGGTG   1298

AACGGTTCTT TCATTTGCTT TGGAAGTTTT AATCTCTAAC AGTGGACAAA GTTACCAGTG   1358

CCTTAAACTC TGTTACACTT TTTGGAAAAA ACTTTGTAGT ATGATAGGTT ATTTTGATGT   1418

AAAGATGTTC TGGATACCAT TATATGTTCC CCCTGTTTCA GAGGCTCAGA TAATATGTAA   1478

ATGGTATGTC ATTCGCTACT ATGATTTAAT TTGAAATATG GTCTTTTGGT TATGAATACT   1538

TTGCAGCACA GCTGAAGGCT GTCTGTTGTA TTCATTGTGG TCATAGCACC TAACAACATT   1598

GTAGCCTCAA TCGAGTGAGA CAGACTAGAA GTTCCTAGTC TTATGATAGC AAATGGCCTC   1658

ATGTCAAATA TTAGATGTAA TTTTGTGTAA GAAATACAGA CTGGATGTAC CACCAACTAC   1718

TACTAATGAC AGGCCTGTCC AACACATCTC CCTTTTCCAT GCTGTGGTAG CCAGCATCGG   1778
```

AAAGAACGCT GATTTAAAGA GGTGAGCGGA ATTTTATTGA CACAGTACCA TTTAATGGGG 1838

AGACAAAAAT GGGGGCCAGG GGAGGGAGAA GTTTCTGTCG TTAAAAACGA GGGAAAGACT 1898

GGACTCTAAA TTCTGTTGAT TAAAGATGAG CTTTGTCTAC CTTCAAAAGT TTGTTTGCTT 1958

ACCCCCTTCA GCCTCTTTTT TAAGTGAAAA TATAACTAAT AACATGTGAA AAGAATAGAA 2018

GCTAAGGTTT AGATAAATAT TGAGCAGATC TATAGGAAGG AACCTGAATA TTGCCATTAT 2078

GCTTGACATG GTTTCCAAAA AATGGTACTC CACATACTTC AGTGAGGGTA AGTATTTTCC 2138

TGTCAAGAAT AGCATTGTAA AAGCATTTTG TAATAATAAA GAATAGCTTT AATGATATGC 2198

TTGTAACTAA AATAATTTTG TAATGTAAAT ACATTTAAAA CATTAAAATA TAATCTCTAT 2258

AAT 2261

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	310
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGGCGCCAGG GTCTTTGTGG ATTGCATGTT GACATTGACC GTGAGATTCG GCTTCAAACC 60

AATACTGCCT TTGGAATATG ACAGAATCAA TAGCCCAGAG AGCTTGTCAA AGACAGATCT 120

ACGGTCATCC TTAACCAAGG CACTTCTTAA GCAGAAAATA TTGTTGAGGT TACCTTTGCT 180

GCTAAAGATC CAATCTTCTA ACGCCACAAC AGCATAGCAA ATCCTAGGAT AATTCACCTC 240

CTCATTTGAC AAACAGAGCT GTAATTCACT TAACAAATTA CGCATTTCTA TCACGTTCAC 300

TACAGCTTAT 310

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	111
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGCACGAGGA AGTATATGGG TATCCGAATG ATGTCACTCA CGTCCAGCAA AGCCAAAGAG 60

CTGAAGGACC GGCACCGGGA CTTCCCAGAC GTGGATCTCA GGAGCGTATA T 111

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	245
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CGAGCTGGAC CCCAAAGAAT ACAGTGGATG GAAAGACAAA CTGCACAGGC AGATGTTTGC 60

CTCATAACAG TCGTAAGTGG AGTCCTGGAA TTTGGACAAG TGCTGTTGGG ATATAGCAAC 120

TTATCTTTGA GTAATGTGAC TAAAGGAAAA AACTTTGACT TTGCCCAGGC ATGAAATTCT 180

TCCTAATGTC AGAACACAGT GCAACCCAGT CACACTGTGG CCAGTAAAAT ACTATTGCCT 240

CATAT 245

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	340
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGAGCTGGTC ATGTGGTTGG CACTAGACTG GTGGCAGGGG CTTCTAGCTG ACTCGCACAG 60

GGATTCTCAC AATAGCCGAC ATCAGAATTT GTGTTGAAGG AACTTGTCTC TTCATCTAAT 120

ATGATAGCGG GAAAAGGAGA GGAAACTACT GCCTTTAGAA AATATAAGTA AAGTGATTAA 180

AATGCTCACG TTACCTTGAC ACATAGTTTT TCAGTCTATG GGTTTAGTTA CTTTACATGG 240

CAAGCATGTA ACTTATATTA ATAGTAATTT GTAAAGTTGG TTGGATAAGC TATCCCTGTT 300

TGCCGGTTCA TGGATTACTT CTCTATAAAA AATATATAT 340

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	76
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AAAAGGCACA TGGTCGTGAT GAAGTTATTA AAGTGGGTGA TACTGTGTGT TTCTTGGTAA 60

ATCCAGTCAG GTAACT 76

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	111
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTTGCTTTGA GAGTGTTAGA CGAACCAGAG GGACACACAG TTTTGACGGT CTTTGGAGGA 60

CCGTTCAACA CCACCACTAC GTGACGATAT CGGTTTTAAC CGTTCGTCGT T 111

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	362
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TGGGCTAAGT AATTTAACTG GGTGTTTATA AAAGTAAAAG GCCAACATTT AATTATTTTG 60

CAAAGCAACC TAAGAGCTAA AGATGTAATT TTTCTTGCAA ATTGTAAATC TTTTGTGTCT 120

CTGAAGACTT CCCTTAAAAT TAGCTCTCTG AGTGAAAAAT CAAAAGAGAC AAAAGACATC 180

TTCGAATCCA TATTTCAAGC CTGGTAGAAT TGGCTTTTCT AGCAGAACCT TTCCAAAAGT 240

TTTATATTGA GATTCATAAC AACACCAAGA ATTGATTTTG TAGCCAACAT TCATTCAATC 300

AGTTATATCA GAGGAGTAGG AGAGAGGAAA CATTTGACTT ATCTGGAAAA GCAAATGTAC 360

TT 362

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	341
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGCACGCGTT TCAGCACACT GAGTTGGGAA TTTCTTATCC CAGAAGACCA ACCAATTTCA 60

TATTTATTTA AGATTGATTC CATCCCCCGT TTTCAAGGAG AATCCCTGCA GTCTCCTTAA 120

AGGTAGAACA AATACTTCTA TTTTTTTTTC ACCATTGTGG GATTGGACTT TAAGAGGTGA 180

CTCTAAAAAA ACAGAGAACA AATATAGTGT CAGTTGTATT AAGCACGGAC CCATATATCA 240

TATTCCACTT AAAAAAATTG CAATTTCCTG TTGCACCTTT TGGCAACTTC TCTTTTCAAT 300

GTAGGGAAAA ACTTAGTCAC CCTGAAAACC CACAAAATAA A 341

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	321
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTCATCGCTG GGATGCTGGT TCTAGAGGCA GCTGTCACGG GAGTTCCTGT TAAAGGTCAA 60

GACCCTGTCA AAGGCCGTGT TCCATCAATG GACAAGATCC CGTTAAAGGA CAAGTTTCAG 120

TTAAAGGTCA AGATAAAGTC AAAGCGCAAG AGCCAGTCAA AGGTCCAGTC TCCACTAAGC 180

CTGGCTCCTG CCCCATTATC TTGATCCGGT GCGCCATGTT GAATCCTCCT AACCGCTGCT 240

TGAAAGATAC TGACTGCCCA GGAATCAAGA AGTGCTGTGA AGGCTCTTGC GGGATGGCCT 300

GTTTCGTTCC CAGTGAGAGG G 321

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	4328
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	single
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GCCAGGTGAA GACCAACCCT GAGGAGAAGA AGTGCTTTGA CCTTATTTCA CATGACAGAA 60

CTTACCACTT TCAAGCTGAA GATGAACAGG AATGTCAAAT ATGGATGTCT GTGCTGCAAA 120

ATAGCAAAGA AGAAGCTTTA AACAATGCAT TTAAGGGGGA TGACAATACT GGAGAAAATA 180

ACATCGTCCA AGAACTGACA AAGGAGATCA TCTCAGAAGT GCAGAGGATG ACGGGCAATG 240

ACGTCTGCTG TGACTGTGGG GCGCCAGATC CTACATGGCT TTCCACCAAC CTGGGCATCC 300

TGACCTGCAT CGAGTGTTCC GGAATCCACC GAGAGCTGGG GGTTCATTAT TCCAGGATGC 360

AGTCCCTGAC CTTAGATGTA CTGGGAACAT CTGAGCTGCT GCTCGCCAAG AATATTGGGA 420

ATGCAGGCTT TAATGAGATC ATGGAATGTT GCCTACCAGC TGAGGACTCA GTCAAACCCA 480

ACCCAGGCAG CGACATGAAT GCAAGAAAGG ACTACATCAC AGCCAAGTAC ATCGAGAGGA 540

GATACGCAAG GAAGAAGCAC GCGGATAACG CGGCGAAGCT TCACAGTCTT TGCGAGGCCG 600

TCAAAACGAG AGATATTTTT GGATTGCTCC AAGCTTATGC TGATGGTGTG GATCTTACGG 660

AAAAAATCCC ACTGGCCAAC GGACATGAGC CGGATGAAAC GGCCCTCCAC CTTGCAGTCA 720

GATCCGTGGA TCGAACCTCT CTTCACATTG TAGACTTTTT AGTTCAGAAC AGTGGGAACC 780

TGGATAAACA GACAGGGAAA GGCAGCACAG CCCTGCACTA CTGCTGCCTG ACCGACAATG 840

CCGAGTGCCT CAAGTTGCTC CTGCGGGGGA AGGCCTCCAT CGAGATAGCA AATGAGTCAG 900

GAGAGACTCC GCTGGACATT GCCAAGCGCC TCAAGCACGA GCACTGTGAG GAGCTGCTGA 960

CCCAAGCCTT ATCTGGAAGA TTTAATTCTC ACGTTCACGT TGAATATGAA TGGCGACTAC 1020

TCCACGAAGA CCTGGATGAA AGTGATGACG ACATGGATGA GAAATTGCAG CCCAGTCCCA 1080

ACCGGCGGGA AGACCGGCCC ATCAGCTTCT ACCAGCTGGG CTCCAACCAG CTTCAGTCTA 1140

ACGCTGTATC TTTGGCCAGA GATGCTGCAA ACCTTGCCAA GGACAAGCAG AGGGCTTTCA 1200

TGCCCAGCAT CTTGCAGAAT GAGACTTACG GAGCCCTCCT GAGTGGCAGC CCACCTCCCG 1260

CCCAGCCTGC AGCCCCCAGC ACCACCAGCG CCCCCCCCGC TTCCTCCACG GAATGTTGGC 1320

AAAGTTCAGA CAGCCTCCTC TGCTAACACC CTGTGGAAGA CAAACTCTGT AAGTGTGGAC 1380

GGTGGAAGCC GGCAGCGATC TTCGTCAGAT CCGCCAGCTG TCCATCCACC GCTGCCCCCT 1440

CTTCGCGTGA CATCTACCAA TCCCCTGACC CCCACGCCGC CCCCACCCGT TGCCAAGACG 1500

CCCAGCGTAA TGGAAGCCTT GAGCCAGCCG AGCAAGCCTG CCCCGCCTGG GATCTCACAG 1560

ATCAGGCCCC CACCTCTGCC CCCACAGCCG CCCAGCCGCC TCCCGCAGAA GAAGCCTGCG 1620

CCAGGGGCTG ACAAGTCCAC CCCACTGACC AACAAAGGCC AACCGAGAGG ACCTGTGGAT 1680

CTCTCTGCAA CGGAAGCTCT GGGTCCTCTG TCCAATGCTA TGGTCCTGCA GCCCCCTGCA 1740

CCCATGCCTA GGAAGTCGCA GGCAACCAAG TTGAAGCCTA AGCGGGTGAA AGCGCTCTAT 1800

AACTGTGTGG CTGACAACCC CGATGAGCTC ACCTTCTCCG AGGGGGATGT GATCATCGTG 1860

GACGGGGAGG AGGACCAGGA GTGGTGGATT GGCCACATTG ATGGAGATCC TGGTCGCAAA 1920

GGCGCATTCC CGGTGTCATT TGTGCACTTT ATCGCTGACT GAATTGCTAC TGAACAAAAG 1980

CATTAACAGT TATGTTCCTG TTTCGTTATT GGTACCAAAA CTCTTGCCAG ATAACCAGTT 2040

TCATGAACTG TTTGTATGGC AGCCCATGTT CTCTAATGCC ACTGCTCTGT TTTAAAAACT 2100

CAGAGGCAAT TTTTACATAT CAGTAATTGT TTTTATAATT TGCATGGTTT TCATGAAACA 2160

TTGCTATGCA TTTATTAGGA AAAACTGAAT TTCCCAACAG GTGAACTGAA AAGTTATTTT 2220

AACTATTATA CATAATCAGA AAGATCCTGC CTCTACGGAA TTAGCTAAAC CTAAAAATGT 2280

TTGCATTAAT GAATAAATTC TTCCTGCATT CCTTGGCCCA GTTCTGGAGT TGGTGACCTT 2340

TATCACAATT ATATTTTAGG CGGCCAGTGA ACTGCTGCTT CAGAAGTCCA TAGCCCAGCT 2400

CTGAACTTTC TCGATAAATG CCATCAGTTC ACCTTTAAAG ACACACATTC CTTTGAAATC 2460

CACCCAGTGT TTAAAAAGCA ACTTGGAAAT TTACACATTA GCATTGTACT TTCTAGCCCT 2520

AATTTGTGAG GTTGCAGCTA TCATTATATT CTGCATGTAT GTATAACCTG TTGTGAACAA 2580

TCATACTTAA CAAAACTACT GATGGTTTAT GACAACGTAG GGTAACTACA GTTCATTCTG 2640

TTCCAGGTTA TATAAAACTG CATTTCCTGA ATTTGGTTAA AAACTAAGGA TGATGGATTC 2700

GAAAACAGTC TTTTAAATTA GTTTATATGC TTTAGGTGTT TTGGAATTTG CCTTCTTGAA 2760

CTTCCTGAGT CACACAGAAA GCAACTGTAC ACAGTAGAAT TCTGTGGCGC AGACCATGCT 2820

GTATTAACAC ATCACTTGCT GTTTCCTACT GAGTGTACCA CTGCCTTCCC TTCTAGCCCA 2880

GGAGAATGTT TACTCAGTTT AGTGTCTTGT ATTTCTATAA TACACCAACA GGAATGGTAG 2940

TCACACTGTC TTGAAATTGA ATCTGTCCAT CTGTTTATAA TCAAGAACAT ATCAGAAATA 3000

TATAGGTCCC AGGTAATACT CCCAAACATC CCACTTTTTA CTGTTTCAGG CCATCATATC 3060

ATTCTTAAGC TACTTGGGGT GGTAGTAGAG GATTAGGTTG TCTATTATAA AACCAAAACT 3120

CATTCGTTTA ATGAACTTGA CTGTCATACC TCTATTTAGT AATTGCGAGG GTAAGATTCA 3180

TAGTAGGAAT ATTGGAAATT TTGGCACTCT GAGAATAAAT AGGCATATGA TACCCACTTG 3240

GACTTTTAAC AAAAGTAAAG GAATAAATTT GCATATAGGT TTGGAAAGTG AGGCAGCAAT 3300

GCTGTTAACT GCATTTGTTG TGATGGTGCA TTTGATTGAA GCAGCTTGTC TTTATTATGC 3360

AAGACTGTGT AGAGTTTTTT TTTTTTTGGC ATTGTACTTT TTGTTTTTGT TATAAGGAAG 3420

ACAGAACAAA CTGGAATGTT TTATGATGTT GTATAGCAAT CGCTTTTTAC CTTTCAAAGT 3480

TCCGGGTAAA AATGTGTTAT ATCTGTAGTT TTTTGTTTTT GTTTTTTTTT AAAGCACTAC 3540

ATCTGTTTTC ACTAATTGTT AATTTCTGTT TGAACCCTTC ATTTAATTTT CTCATAGATT 3600

TAAGTAAACA AGGATGTATT TTGCACACGC TCGCACTTAT GTCTATTTTA ACAATCTCCT 3660

GCATCTGTAT TTTATAGTCA GCCTTTTGAC CACCTGGTGC CAGCTATATA AGGAATAAAG 3720

TTGATTCATA TCAACATTAG AACTCCAGTC CCAAACTAAT CTGTCAGGTT CACTGGTACA 3780

TAAATACCTA GGAAATATTT TTCCAGTCTA CATTTGGTGC TATGTGCAGT AACTAATAGT 3840

ACTCTTACCA GAGGAGAAAT TATATAACGA CCCTGCTAAT ATCTTTCTTA GTTATTTGCT 3900

CCTTCAAATT AAAAAAGCAA CTAAGAGAAA GAAAAACATT GTAGATATCT ATTTATATTT 3960

AAAGTTTATG AAACATGAAC TGCAGCTGCA GGATTCTGGC ATTTTGCATG CCATTCTCCA 4020

TCAGATCTGG GATGATGGCT CAGAACATGT ACACAGACTA AGAGTAACTG TGTGATCTGT 4080

TAAGGGGTGG ATAACATAAT ATGCAGCTTA GGATGCTATT TTGAGATGTA TGATATCAGT 4140

TCATTCACCT GATTACTTTG GTTGCAGCAC AACTGTATAT ATTGTATAAC CGAAATTGAT 4200

TATTTTCATT GTCCTTATGC AGTGATTTAT AATTAGAGCA TGTTTAATAA GTTTACTATT 4260

CTTGTTAACT AGTCATTTGA CTGGAAAAAA ATAAAATACT TTTAAATGGA AAAAAAAAAA 4320

AAAAAAAA 4328

Claims

1. A method for determining the presence of cancerous cells in a tissue from a patient, which method comprises the steps of:

providing a nucleic acid probe comprising a nucleotide sequence at least 8 nucleotides in length which is identical to a portion of the coding sequence of a candidate tumor suppressor gene;

obtaining from a patient a first tissue sample potentially comprising cancerous cells;

providing a second tissue sample, substantially all of the cells which are non-cancerous;

contacting said nucleic acid under stringent hybridizing conditions with RNA of each of said first and second tissue samples; and

comparing (a) the amount of hybridization of said nucleic acid probe with said RNA of said first tissue sample, with (b) the amount of hybridization of said nucleic acid probe with said RNA of said second tissue sample, wherein an amount of hybridization with said RNA of said first tissue sample less than one third the amount of hybridization with said RNA of said second tissue sample indicates the presence of cancerous cells in said first tissue sample.

2. A method for determining the presence of cancerous cells in a tissue from a patient, which method comprises the steps of:

providing an antibody specific for the gene product of a candidate tumor suppressor gene, said gene product being present in cancerous tissue of a given tissue type at a level less than one third the level of said gene product in noncancerous tissue of said given tissue type;

obtaining from a patient a first sample of tissue of said given tissue type, said first sample potentially comprising cancerous cells;

providing a second sample of tissue of said given tissue type, essentially all of the cells of which sample are non-cancerous;

contacting said antibody with protein of said first and second samples under conditions permitting immunocomplex formation; and

comparing (a) the amount of immunocomplex formation in said first sample, with (b) the amount of immunocomplex formation in said second sample, wherein an amount of immunocomplex formation in said first sample less than one third the amount of immunocomplex formation in said second sample indicates the presence of cancerous cells in said first sample of tissue.

3. The method of claim 1 or 2 wherein said candidate tumor suppressor gene is chosen from a gene encoding keratin 5, NB-1 gene product, fibronectin, connexin 26, glutathione-S-transferase pi, CaN19 protein, small proline-rich (spr-1) protein, amphiregulin, thymosin beta-4, gamma actin, calpactin light chain (p11), HBp17, myosin regulatory light chain, v-fos transformation effector protein, or one of the following mitochondrial genome-encoded proteins: URF4, Co III, or ATPase.

4. The method of claim 1 or 2, wherein said candidate tumor suppressor gene comprises a sequence which hybridizes under stringent conditions to a sequence shown in SEQ ID NO: 3 (U1), SEQ ID NO: 4 (U2), SEQ ID NO: 5 (U3), SEQ ID NO: 6 (U4), SEQ ID NO: 7 (U5), SEQ ID NO: 8 (U6), SEQ ID NO: 9 (U7), SEQ ID NO: 10 (U8), SEQ ID NO: 11 (U9), or SEQ ID NO: 12 (U10).

5. The method of claim 1 wherein said contacting step comprises performing a northern analysis or an in situ hybridization analysis.

6. A method for identifying a drug useful for treatment of a cancer, comprising the steps of:

identifying a candidate class II tumor suppressor gene, expression of which is suppressed in a given type of cancer cell;

providing a first and a second sample of said given type of cancer cell;

determining the level of expression of said gene in said first sample;

treating said second sample with a candidate drug; and

determining the level of expression of said gene in said second sample after treatment with said candidate drug, wherein a drug which increases the level of expression of said gene in said second sample, compared to the level of expression of said gene in said untreated first sample, is potentially useful for treatment of said given type of cancer cell.

7. The method of claim 6, wherein said drug alters a signal transduction pathway in said given type of cancer cell.

8. The method of claim 6, wherein said drug increases synthesis or processing of the mRNA of said candidate tumor suppressor gene in said given type of cancer cell.

9. The method of claim 6, wherein said drug decreases degradation of the messenger RNA transcribed

from said candidate tumor suppressor gene in said given type of cancer cell.

10. The method of claim 6, wherein said drug increases protein synthesis from mRNA transcribed from said candidate tumor suppressor gene in said given type of cancer cell.

11. The method of claim 6, wherein said drug decreases degradation of the gene product of said candidate tumor suppressor gene in said given type of cancer cell.

12. The method of claim 6, wherein said drug demethylates methylated residues on DNA.

13. A method for cancer cell treatment comprising the steps of:

identifying a patient with cancer cells characterized by a low level of expression of a candidate class II tumor suppressor gene, compared to the level of expression of said gene in normal cells; and

treating said patient with a compound which raises the level of expression of said gene in said cancer cells.

14. The method of claim 13 wherein said gene encodes keratin 5, NB-1 gene product, fibronectin, connexin 26, glutathione-S-transferase pi, CaN19 protein, small proline-rich (spr-1) protein, amphiregulin, thymosin beta-4, gamma actin, calpactin light chain (p11), HBp17, myosin regulatory light chain, v-fos transformation effector protein, or one of the following mitochondrial genome-encoded proteins: URF4, Co III, or ATPase6.

15. The method of claim 13, wherein said candidate tumor suppressor gene comprises a sequence which hybridizes under stringent conditions to a sequence shown in SEQ ID NO: 3 (U1), SEQ ID NO: 4 (U2), SEQ ID NO: 5 (U3), SEQ ID NO: 6 (U4), SEQ ID NO: 7 (U5), SEQ ID NO: 8 (U6), SEQ ID NO: 9 (U7), SEQ ID NO: 10 (U8), SEQ ID NO: 11 (U9), or SEQ ID NO: 12 (U10).

16. The method of claim 13, wherein said compound demethylates methylated residues in DNA.

17. A method for cancer cell treatment comprising the steps of:

identifying a patient with a cancer cell characterized by a low level of expression of a candidate tumor suppressor gene, compared to the level of expression of said gene in normal cells of the same tissue type as said cancer cells; and

introducing into said cancer cell a nucleic acid encoding a candidate tumor suppressor gene.

18. The method of claim 17, wherein said nucleic acid includes an expression control element permitting expression of said candidate tumor suppressor gene in said cancer cell.

19. The method of claim 17, wherein said candidate tumor suppressor gene encodes keratin 5, NB-1 gene product, fibronectin, connexin 26, glutathione-S-transferase pi, CaN19 protein, spr-1 protein, amphiregulin, thymosin beta-4, gamma actin, calpactin light chain (p11), HBp17, myosin regulatory light chain, v-fos transformation effector protein, or one of the

following mitochondrial genome-encoded proteins: URF4, Co III, or ATPase6.

20. The method of claim 17, wherein said candidate tumor suppressor gene comprises a sequence which hybridizes under stringent conditions to a sequence shown in SEQ ID NO: 3 (U1), SEQ ID NO: 4 (U2), SEQ ID NO: 5 (U3), SEQ ID NO: 6 (U4), SEQ ID NO: 7 (U5), SEQ ID NO: 8 (U6), SEQ ID NO: 9 (U7), SEQ ID NO: 10 (U8), SEQ ID NO: 11 (U9), or SEQ ID NO: 12 (U10).

21. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 3 (U1).

22. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 4 (U2).

23. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 5 (U3).

24. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 6 (U4).

25. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 7 (U5).

26. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 8 (U6).

27. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 9 (U7).

28. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 10 (U8).

29. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 11 (U9).

30. An isolated DNA which hybridizes under stringent conditions to a DNA having the sequence shown in SEQ ID NO: 12 (U10).

31. A method of determining the level of a candidate tumor suppressor gene product in a biological fluid, said method comprising

obtaining a sample of a biological fluid from a person;

contacting proteins in said sample with an antibody specific for a candidate tumor suppressor gene product; and

determining the amount of immune complex formation by said antibody, said amount being indicative of the level of said gene product in said sample.

32. The method of claim 31, wherein said biological fluid is blood, urine, or saliva.

33. The method of claim 31, wherein said amount is compared to the amount of immune complex formation by said antibody in a normal control sample.

34. The method of claim 31, wherein said amount is compared to the amount of immune complex formation by said antibody in a sample previously or subsequently obtained from said person.

FIGURE

```
                                                                                GATTTAATCCTATGACAAACTAAGTT -175
GGTTCTGTCTTCACCTGTTTTGGTGAGGTTGTGTAAGAGTTGGTGTTTGCTCAGGAAGAGATTTAAGCATGCTTGCTTACCCAGACT  -88
CAGAGAAGTCTCCCTGTTCTGTCCTAGCTAGTGATTCCTGTGTTGTGTGCATTCGTCTTTTCCAGAGCAAACCGCCCAGAGTAGAAG   -1

    ATG GAT TGG GGC ACG CTG CAG ACG ATC CTG GGG GGT GTG AAC AAA CAC TCC ACC AGC ATT GGA AAG   66
  1  M  D  W  G  T  L  Q  T  I  L  G  G  V  N  K  H  S  T  S  I  G  K

    ATC TGG CTC ACC GTC CTC TTC ATT TTT CGC ATT ATG ATC CTC GTT GTG GCT GCA AAG GAG GTG TGG  132
 23  I  W  L  T  V  L  F  I  F  R  I  M  I  L  V  V  A  A  K  E  V  W

    GGA GAT GAG CAG GCC GAC TTT GTC TGC AAC ACC CTG CAG CCA GGC TGC AAG AAC GTG TGC TAC GAT  198
 45  G  D  E  Q  A  D  F  V  C  N  T  L  Q  P  G  C  K  N  V  C  Y  D

    CAC TAC TTC CCC ATC TCC CAC ATC CGG CTA TGG GCC CTG CAG CTG ATC TTC GTG TCC AGC CCA GCG  264
 67  H  Y  F  P  I  S  H  I  R  L  W  A  L  Q  L  I  F  V  S  S  P  A

    CTC CTA GTG GCC ATG CAC GTG GCC TAC CGG AGA CAT GAG AAG AAG AGG AAG TTC ATC AAG GGG GAG  330
 89  L  L  V  A  M  H  V  A  Y  R  R  H  E  K  K  R  K  F  I  K  G  E

    ATA AAG AGT GAA TTT AAG GAC ATC GAG GAG ATC AAA ACC CAG AAG GTC CGC ATC GAA GGC TCC CTG  396
111  I  K  S  E  F  K  D  I  E  E  I  K  T  Q  K  V  R  I  E  G  S  L

    TGG TGG ACC TAC ACA AGC AGC ATC TTC TTC CGG GTC ATC TTC GAA GCC GCC TTC ATG TAC GTC TTC  462
133  W  W  T  Y  T  S  S  I  F  F  R  V  I  F  E  A  A  F  M  Y  V  F

    TAT GTC ATG TAC GAC GGC TTC TCC ATG CAG CGG CTG GTG AAG TGC AAC GCC TGG CCT TGT CCC AAC  528
155  Y  V  M  Y  D  G  F  S  M  Q  R  L  V  K  C  N  A  W  P  C  P  N

    ACT GTG GAC TGC TTT GTG TCC CGG CCC ACG GAG AAG ACT GTC TTC ACA GTG TTC ATG ATT GCA GTG  594
177  T  V  D  C  F  V  S  R  P  T  E  K  T  V  F  T  V  F  M  I  A  V

    TCT GGA ATT TGC ATC CTG CTG AAT GTC ACT GAA TTG TGT TAT TTG CTA ATT AGA TAT TGT TCT GGG  660
199  S  G  I  C  I  L  L  N  V  T  E  L  C  Y  L  L  I  R  Y  C  S  G

    AAG TCA AAA AAG CCA GTT TAA                                                          681
221  K  S  K  K  P  V  -

CGCATTGCCCAGTTGTTAGATTAAGAAATAGACAGCATGAGAGGGATGAGGCAACCCGTGCTCAGCTGTCAAGGCTCAGTCGCCAGC  768
ATTTCCCAACACAAAGATTCTGACCTTAAATGCAACCATTTGAAACCCCTGTAGGCCTCAGGTGAAACTCCAGATGCCACAATGAGC  855
TCTGCTCCCCTAAAGCCTCAAAACAAAGGCCTAATTCTATGCCTGTCTTAATTTTCTTTCACTTAAGTTAGTTCCACTGAGACCCCA  942
GGCTGTTAGGGGTTATTGGTGTAAGGTACTTTCATATTTTAAACAGAGGATATCGGCATTTGTTTCTTTCTCTGAGGACAAGAGAAA 1029
AAAGCCAGGTTCCACAGAGGACACAGAGAAGGTTTGGGTGTCCTCCTGGGGTTCTTTTTGCCAACTTTCCCCACGTTAAAGGTGAAC 1116
ATTGGTTCTTTCATTTGCTTTGGAAGTTTTAATCTCTAACAGTGGACAAAGTTACCAGTGCCTTAAACTCTGTTACACTTTTTGGAA 1203
GTGAAAACTTTGTAGTATGATAGGTTATTTTGATGTAAAGATGTTCTGGATACCATTATATGTTCCCCCTGTTTCAGAGGCTCAGAT 1290
TGTAATATGTAAATGGTATGTCATTCGCTACTATGATTTAATTTGAAATATGGTCTTTTGGTTATGAATACTTTGCAGCACAGCTGA 1377
GAGAGGCTGTCTGTTGTATTCATTGTGGTCATAGCACCTAACAACATTGTAGCCTCAATCGAGTGAGACAGACTAGAAGTTCCTAGT 1464
TGGCTTATGATAGCAAATGGCCTCATGTCAAATATTAGATGTAATTTTGTGTAAGAAATACAGACTGGATGTACCACCAACTACTAC 1551
CTGTAATGACAGGCCTGTCCAACACATCTCCCTTTTCCATGCTGTGGTAGCCAGCATCGGAAAGAACGCTGATTTAAAGAGGTGAGC 1638
TTGGGAATTTTATTGACACAGTACCATTTAATGGGGAGACAAAAATGGGGGCCAGGGGAGGGAGAAGTTTCTGTCGTTAAAAACGAG 1725
TTTGGAAAGACTGGACTCTAAATTCTGTTGATTAAAGATGAGCTTTGTCTACCTTCAAAAGTTTGTTTGCTTACCCCCTTCAGCCTC 1812
CAATTTTTTAAGTGAAAATATAACTAATAACATGTGAAAAGAATAGAAGCTAAGGTTTAGATAAATATTGAGCAGATCTATAGGAAG 1899
ATTGAACCTGAATATTGCCATTATGCTTGACATGGTTTCCAAAAAATGGTACTCCACATACTTCAGTGAGGGTAAGTATTTTCCTGT 1986
TGTCAAGAATAGCATTGTAAAAGCATTTTGTAATAATAAAGAATAGCTTTAATGATATGCTTGTAACTAAAATAATTTTGTAATGTA 2073
TCAAATACATTTAAAACATTAAAATATAATCTCTATAAT                                              2112
```

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/01624

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C07H 21/00; C12N 15/00; C12Q 1/68
US CL : 536/27; 435/6

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	536/27; 435/6

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

APS, BIOSIS
search terms: class II tumo(u)r suppressor? and assay or diagnos? and (nucleic acid or in situ) hybridiz?

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
Y	Science, volume 246, Issued 15 December, 1989, R. Sager, "Tumor Suppressor Genes: The Puzzle and the Promise", pages 1406-1412, see entire document.	21, 1,3,4,5
Y	US, A, 4,888,278 (Singer et al.) 19 December, 1989, see entire document.	21, 1,3,4,5
Y	Virology, Volume 160, issued 1987, R. Cattaneo et al., "Altered ratios of measles virus transcripts in diseased human brains", pages 523-526, see entire document.	21,1,3,4,5

21, 1,3,4,5

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
22 May 1992	10 JUN 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	LORRAINE M. SPECTOR, PH.D.

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

Please See Attached Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

21,1,3,4,5 (Telephone Practice)

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.